describes an interface using video images of hand gestures. A video signal having a frame image containing regions is input to a processor. A plurality of regions in the frame are defined and screened to locate an image of a hand in one of the regions. The hand image is processed to locate extreme curvature values, such as peaks and valleys, corresponding to predetermined hand positions and gestures. The number of peaks and valleys are then used to identify and correlate a predetermined hand gesture to the hand image for effectuating a particular computer operation or function.

[0016] U.S. Pat. No. 6,232,960 to Goldman, entitled "Data Input Device," issued May 15, 2001, describes a data entry device including a plurality of sensing devices worn on a user's fingers, and a flat light-weight keypad for transmitting signals indicative of data entry keyboard functions to a computer or other data entry device. The sensing devices include sensors that are used to detect unique codes appearing on the keys of the keypad or to detect a signal, such as a radar signal, generated by the signal-generating device mounted to the keypad. Pressure sensitive switches, one associated with each finger, contain resistive elements and optionally sound generating means and are electrically connected to the sensors so that when the switches are pressed they activate a respective sensor and also provide a resistive force and sound comparable to keys of a conventional keyboard.

[0017] U.S. Pat. No. 6,115,482, to Sears et al., entitled "Voice Output Reading System with Gesture Based Navigation," issued Sep. 5, 2000, describes an optical-input print reading device with voice output for people with impaired or no vision. The user provides input to the system via hand gestures. Images of the text to be read, on which the user performs finger- and hand-based gestural commands, are input to a computer, which decodes the text images into their symbolic meanings through optical character recognition, and further tracks the location and movement of the hand and fingers in order to interpret the gestural movements into their command meaning. In order to allow the user to select text and align printed material, feedback is provided to the user through audible and tactile means. Through a speech synthesizer, the text is spoken audibly. For users with residual vision, visual feedback of magnified and image enhanced text is provided.

[0018] U.S. Pat. No. 6,204,852, to Kumar et al., entitled "Video Hand Image Three-Dimensional Computer Interface," issued Mar. 20, 2001, describes a video gesture-based three-dimensional computer interface system that uses images of hand gestures to control a computer and that tracks motion of the user's hand or an elongated object or a portion thereof in a three-dimensional coordinate system with five degrees of freedom. During operation of the system, hand images from cameras are continually converted to a digital format and input to a computer for processing. The results of the processing and attempted recognition of each image are then sent to an application or the like executed by the computer for performing various functions or operations. When the computer recognizes a hand gesture as a "point" gesture with one finger extended, the computer uses information derived from the images to track three-dimensional coordinates of the extended finger of the user's hand with five degrees of freedom. The computer utilizes two-dimensional images obtained by each camera to derive three-dimensional position (in an x, y, z coordinate system) and orientation (azimuth and elevation angles) coordinates of the extended finger.

[0019] U.S. Pat. No. 6,002,808, to Freeman, entitled "Hand Gesture Control System," issued Dec. 14, 1999, describes a system for recognizing hand gestures for the control of computer graphics, in which image moment calculations are utilized to determine an overall equivalent rectangle corresponding to hand position, orientation and size, with size in one embodiment correlating to the width of the hand.

[0020] These and other systems use cameras or other light-sensitive sensors to detect user actions to implement virtual keyboards or other input devices. Such systems suffer from some shortcomings that limit both their reliability and the breadth of applications where the systems can be used. First, the time at which a finger touches the surface can be determined only with an accuracy that is limited by the camera's frame rate. For instance, at 30 frames per second, finger landfall can be determined only to within 33 milliseconds, the time that elapses between two consecutive frames. This may be satisfactory for certain applications, but in some cases may introduce an unacceptable delay, for example in the case of a musical instrument.

[0021] A second limitation of such systems is that it is often difficult to distinguish gestures made intentionally for the purpose of communication with the device from involuntary motions, or from motions made for other purposes. For instance, in a virtual keyboard, it is often difficult to distinguish, using images alone whether a particular finger has approached the typing surface in order to strike a virtual key, or merely in order to rest on the typing surface, or perhaps has just moved in sympathy with another finger that was actually striking a virtual key. When striking a virtual key, other fingers of the same hand often move down as well, and because they are usually more relaxed than the finger that is about to strike the key, they can bounce down and come in very close proximity with the typing surface, or even come in contact with it. In a camera-based system, two fingers may be detected touching the surface, and the system cannot tell whether the user intended to strike one key or to strike two keys in rapid succession. In addition, typists often lower their fingers onto the keyboard before they start typing. Given the limited frame rate of a camera-based system, it may be difficult to distinguish such motion of the fingers from a series of intended keystrokes.

[0022] Similarly, another domain in which user actions are often misinterpreted is virtual controls. Television sets, stereophonic audio systems, and other appliances are often operated through remote controls. In a vehicle, the radio, compact disc player, air conditioner, or other device are usually operated through buttons, levers, or other manual actuators. For some of these applications, it may be desirable to replace the remote control or the manual actuators with virtual controls. A virtual control is a sensing mechanism that interprets the gestures of a user in order to achieve essentially the same function of the remote control or manual actuator, but without requiring the user to hold or touch any physical device. It is often difficult for a virtual control device to determine when the user actually intends to communicate with the device.

[0023] For example, a virtual system using popup menus can be used to navigate the controls of a television set in a